



Analisis Sentimen Aplikasi KA Bandara Menggunakan Metode Support Vector Machine dan Random Forest

Reagen Westerdam Sean Jatindra¹, Farhan Syahputra Wiyono², Muhammad Hilmy
Maulana³, Shindi Shella May Wara⁴, Kartika Maulida Hindrayani⁵

^{1,2,3,4,5}Sains Data, UPN "Veteran" Jawa Timur

¹23083010013@student.upnjatim.ac.id

²23083010088@student.upnjatim.ac.id

³23083010106@student.upnjatim.ac.id

⁴shindi.shella.fasikom@upnjatim.ac.id

⁵kartika.maulida.ds@upnjatim.ac.id

Abstract: Digital transformation has significantly impacted public transportation services, including the adoption of the KA Bandara application that facilitates ticket booking and travel information access. User reviews on the Google Play Store serve as a valuable source for analyzing user perceptions. In relation to service quality, this research conducts sentiment analysis on 1,000 reviews of the KA Bandara application using two classification algorithms: Support Vector Machine (SVM) and Random Forest (RF). The review data were collected via web scraping and preprocessed through steps including processes such as case folding, tokenization, stopword elimination, and stemming. Text features were extracted using TF-IDF, and sentiment labels were assigned based on star ratings (positive: 4–5; negative: 1–3). The models were assessed based on metrics including accuracy, precision, recall, and F1-score. The findings indicate that SVM outperforms RF, achieving an accuracy of 89%, while RF recorded 86.5%. This research confirms the superior performance of SVM for sentiment classification of digital app reviews. The findings offer actionable insights for developers to enhance service quality by leveraging user feedback.

Keywords: Sentiment Analysis, TF-IDF, Support Vector Machine, Random Forest, Web Scraping

Abstrak: Transformasi digital telah membawa perubahan besar dalam layanan transportasi publik, salah satunya melalui aplikasi KA Bandara yang memudahkan pengguna dalam memesan tiket dan mengakses informasi perjalanan. Ulasan pengguna di Google Play Store menjadi sumber data penting untuk menganalisis persepsi terhadap kualitas layanan. Penelitian bertujuan untuk menganalisis sentimen masyarakat terhadap aplikasi KA Bandara menggunakan dua algoritma klasifikasi: Support Vector Machine (SVM) dan Random Forest (RF). Data didapatkan dengan teknik web scraping, lalu diproses melalui tahapan pra-pemrosesan teks seperti case folding, tokenisasi, stopword removal, dan stemming. Fitur teks diekstraksi menggunakan metode TF-IDF, dan pelabelan sentimen didasarkan pada rating bintang (positif: 4–5; negatif: 1–3). Evaluasi model dilakukan dengan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa Support Vector Machine unggul dengan akurasi 89%, sementara RF mencatat 86,5%. Penelitian ini menunjukkan bahwa Support Vector Machine lebih efektif untuk klasifikasi sentimen pada ulasan aplikasi digital. Temuan ini dapat menjadi referensi bagi pengembang untuk meningkatkan kualitas layanan berdasarkan opini pengguna.

Kata kunci: Analisis Sentimen, KA Bandara, TF-IDF, Support Vector Machine, Random Forest

I. PENDAHULUAN

Transformasi digital telah mengubah cara masyarakat mengakses layanan publik, termasuk sektor transportasi. Aplikasi KA Bandara merupakan salah satu contoh layanan



SENADA

Seminar Nasional Sains Data 2025 (SENADA 2025)

UPN “Veteran” Jawa Timur

E-ISSN 2808-5841

P-ISSN 2808-7283

digital yang memudahkan masyarakat dalam melakukan pemesanan tiket dan memperoleh informasi perjalanan kereta bandara secara daring. Tingkat adopsi yang tinggi dari pengguna terhadap solusi transportasi digital di Indonesia [1] *KA Bandara* telah diimplementasikan untuk mempermudah akses masyarakat terhadap layanan administratif dan sosial. Analisis sentimen terhadap komentar pengguna pada aplikasi ini menjadi krusial dalam mengevaluasi kinerja dan persepsi masyarakat terhadap layanan digital tersebut [2]

Seiring meningkatnya jumlah pengguna, berbagai opini dan pengalaman terkait kualitas layanan terekam dalam bentuk ulasan teks di platform Google Play Store. Ulasan-ulasan ini menjadi sumber data yang kaya, namun tidak terstruktur. Oleh karena itu, dibutuhkan metode pengolahan bahasa alami yang mampu mengidentifikasi dan mengklasifikasikan opini pengguna secara sistematis. Pendekatan yang sering digunakan adalah analisis sentimen, yaitu proses untuk menentukan polaritas suatu teks apakah bersentimen positif, negatif, atau netral [3].

Dalam penelitian ini, kita melakukan analisis sentimen terhadap ulasan aplikasi *KA Bandara* menggunakan dua metode klasifikasi yaitu Support Vector Machine (SVM) dan Random Forest (RF). SVM dikenal efektif dalam menangani data berdimensi besar dengan margin klasifikasi yang maksimal [4], sedangkan RF merupakan metode berbasis ensemble learning yang robust terhadap overfitting dan mampu menangani data teks yang kompleks [5].

Data ulasan didapat dari web scraping Google Play Store menggunakan pustaka Python. Tahapan praproses teks dilakukan untuk membersihkan dan menormalisasi data, termasuk penghapusan karakter khusus, tokenisasi, dan stopword removal. Selanjutnya, fitur teks direpresentasikan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) yang mempertimbangkan bobot pentingnya suatu kata dalam korpus dokumen [6]. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. terhadap dataset berlabel manual.

Hasil penelitian ini diharapkan dapat memberikan gambaran komparatif performa kedua algoritma dalam klasifikasi sentimen, serta memberikan masukan bagi pengelola aplikasi untuk meningkatkan kualitas layanan berdasarkan persepsi pengguna.



II. METODE PENELITIAN

II.1. Pengumpulan Dataset (Web Scraping)

Data ulasan pengguna aplikasi KA Bandara diambil dari Google Play Store melalui proses web scraping.. Data yang dikumpulkan berupa ulasan teks pengguna dan rating bintang. Pengambilan dilakukan secara otomatis menggunakan pustaka Python google play scraper, lalu disimpan dalam format.csv untuk diproses lebih lanjut.

II.2. Pra-pemrosesan Data

Pra-pemrosesan merupakan tahap krusial dalam analisis sentimen, khususnya ketika berhadapan dengan data ulasan pengguna dari platform Google Play Store. Langkah pertama yang dilakukan pada data teks adalah **Case Folding**, yakni proses konversi seluruh karakter ke dalam bentuk huruf kecil sebagai upaya standarisasi data sebelum tahap analisis.. Langkah kedua **Cleansing**, yaitu penghapusan tanda baca, angka, serta karakter khusus. Langkah ketiga adalah **Tokenisasi**, yaitu memecah teks ulasan menjadi satuan kata atau token. Langkah keempat adalah **Stopword**, yaitu bertujuan untuk menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis data. Tahap yang terakhir adalah **Stemming**, yaitu mengembalikan setiap kata ke bentuk dasarnya. [5]

II.3. Ekstraksi Fitur

Dalam tahap ini, fitur teks dikonversi menjadi nilai numerik melalui metode Term Frequency-Inverse Document Frequency (TF-IDF), yang berfungsi memberikan bobot terhadap kata-kata penting berdasarkan frekuensinya dalam sebuah dokumen relatif terhadap keseluruhan dokumen.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t)$$

dengan:

$$\text{TF}(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}}, \quad \text{IDF}(t) = \log \left(\frac{N}{1 + |\{d : t \in d\}|} \right)$$

di mana:

- $f_{t,d}$ menunjukkan seberapa sering kata t muncul dalam dokumen d ,
- N adalah jumlah total dokumen dalam dataset.
- $|\{d : t \in d\}|$ adalah jumlah dokumen yang mengandung kata t



II.4. Labeling Data

Data ulasan diklasifikasikan ke dalam dua label sentimen, yaitu positif dan negatif, berdasarkan nilai rating:

- Rating 4–5: positif
- Rating 1–3: negatif

Label tersebut digunakan sebagai variabel target pada tahap pelatihan model klasifikasi.

II.5. Model Klasifikasi

Dua algoritma klasifikasi yang digunakan untuk menganalisis sentimen adalah:

- Model Support Vector Machine (SVM)
SVM berfungsi untuk membangun hyperplane pemisah yang optimal, dengan margin selebar mungkin antara dua kelompok data yang berbeda kelas. Fungsi objektifnya dinyatakan sebagai:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{dengan syarat} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad \forall i$$

Dalam kasus non-linear, digunakan kernel seperti Radial Basis Function (RBF):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

- Model Random Forest (RF)
Random Forest adalah metode ensemble yang terdiri dari sejumlah pohon keputusan, di mana hasil prediksi ditentukan berdasarkan voting mayoritas dari seluruh pohon yang dihasilkan.. Proses pembentukan pohon dilakukan secara acak baik dari sisi fitur maupun sampel data (*bagging*), sehingga mengurangi risiko *overfitting*. Prediksi akhir diperoleh dari voting mayoritas:

$$\hat{y} = \text{mode}(\{h_1(x), h_2(x), \dots, h_n(x)\})$$

di mana $h_i(x)$ adalah prediksi dari pohon ke- i .

II.6. Evaluasi Model

Model dikaji performanya melalui empat metrik evaluasi yang menjadi standar dalam pengukuran kinerja algoritma klasifikasi.

- Akurasi:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$



- Presisi:

$$\text{Presisi} = \frac{TP}{TP + FP}$$

- Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- F1-Score:

$$F1 = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$$

Model diuji menggunakan data validasi terpisah dengan metode hold-out (80% data latih, 20% data uji).

III. HASIL DAN PEMBAHASAN

III.1. Statistik Data Ulasan

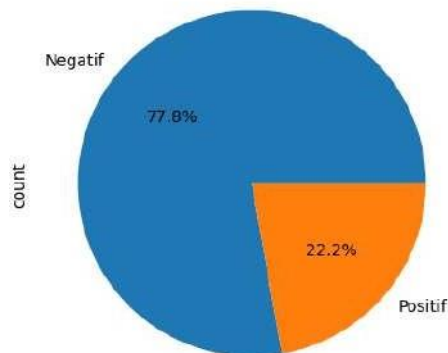
Data yang diperoleh dari Google Play Store melalui proses scraping web yang mencakup 1.000 ulasan pengguna terhadap aplikasi KA Bandara. Setiap ulasan disertai dengan rating bintang dari 1 hingga 5. Labelisasi data dilakukan berdasarkan ketentuan:

- Rating 4–5 → Sentimen Positif
- Rating 1–3 → Sentimen Negatif

Tabel 1. Distribusi Pelabelan

Sentimen	Jumlah Ulasan
Positif	680
Negatif	320

Proporsi Komentar Positif vs Negatif





Tabel 3. Hasil Evaluasi Model

Algoritma	Akurasi	Presisi	Recall	F1-Score
SVM	0.89	0.895	0.876	0.885
RF	0.865	0.872	0.841	0.856

III.5. Analisis dan Diskusi

Dari hasil evaluasi, algoritma SVM menunjukkan kinerja terbaik pada semua metrik klasifikasi. Ini membuktikan efektivitas SVM dalam menangani data teks berdimensi tinggi yang direpresentasikan dengan TF-IDF. Sebaliknya, RF tetap menunjukkan performa yang baik namun sedikit lebih rendah.

Untuk memperkuat analisis sentimen, dibuat visualisasi WordCloud terpisah untuk masing-masing label:



Gambar 4a.



Gambar 4b

Gambar 4. WordCloud : (4a) WordCloud Positif (4b) WordCloud Negatif

kata-kata seperti “bagus”, “mudah”, “membantu”, dan “Good” mendominasi komentar positif, mencerminkan kepuasan terhadap kemudahan penggunaan aplikasi. Sebaliknya, Gambar 4b menunjukkan kata-kata seperti “error”, “login”, “refund”, dan “tiket” sebagai representasi dominan dari keluhan pengguna, menunjukkan masalah teknis dan proses pembayaran.

IV. KESIMPULAN

Penelitian ini berhasil melakukan analisis sentimen terhadap 3.567 Ulasan pengguna aplikasi KA Bandara diperoleh melalui proses scraping dari Google Play Store. Proses pelabelan



didasarkan pada rating ulasan, dengan pembagian menjadi dua kategori: Positif (rating 4-5) dan negatif (rating 1-3)[8]

Data ulasan KA Bandara hasil scraping diproses menggunakan *preprocessing* teks yang mencakup case folding, cleansing, tokenisasi, stopword removal, dan stemming. Fitur teks kemudian diekstraksi menggunakan metode TF-IDF untuk menghasilkan representasi numerik yang dapat dimasukkan ke dalam model klasifikasi.

Dua algoritma klasifikasi diuji, yaitu Support Vector Machine (SVM) dan Random Forest (RF), menggunakan evaluasi hold - out dengan 80% data latih dan 20% data uji. Hasil evaluasi menunjukkan bahwa algoritma SVM memberikan kinerja terbaik dengan akurasi 89%, presisi 89,5%, recall 87,6%, dan F1-score 88,5%. Sementara itu, Random Forest memberikan hasil yang cukup baik namun sedikit lebih rendah.[9]

Dengan demikian, dapat disimpulkan bahwa metode Support Vector Machine lebih unggul dibandingkan Rando Forest dalam melakukan klasifikasi sentimen terhadap data ulasan aplikasi KA Bandara. Penelitian ini memberikan kontribusi praktis bagi pengembangan aplikasi dalam memahami persepsi pengguna serta meningkatkan kualitas layanan berdasarkan umpan balik yang dianalisis secara otomatis[10]

UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dalam proses penyusunan penelitian ini. Ucapan terima kasih juga ditujukan kepada Universitas Pembangunan Nasional "Veteran" Jawa Timur, khususnya Program Studi Sains Data, atas fasilitas dan kesempatan yang diberikan dalam pelaksanaan penelitian. Selain itu, penulis menghargai kontribusi dari Pemerintah Provinsi Jawa Barat serta Jabar Digital Service atas pengembangan aplikasi Sapawarga yang menjadi fokus kajian dalam studi ini. Berkat kerja sama dan dukungan berbagai pihak, penelitian ini dapat terselesaikan dengan baik dan memberikan kontribusi akademik dalam pengembangan layanan publik berbasis data.

REFERENSI

1. Purwarianti, A., & Crisdayanti, I. (2019). *Aspect-based sentiment analysis on Indonesian hotel reviews using neural network classifiers*. *Procedia Computer Science*, 157, 465-472. <https://doi.org/10.1016/j.procs.2019.09.006>
2. Wara, S. S. M., Adziima, A. F., Pratama, A. R., & Nasrudin, M. (2024, November). *Predictive analysis of government application comment on Playstore with Clustered Support Vector Machine*. Paper presented at the 2024 IEEE 10th Information Technology International Seminar (ITIS), Surabaya, Indonesia.



3. Sari, R. F., & Wibowo, A. (2018). *Sentiment Analysis on Customer Review Using Multinomial Naive Bayes and Support Vector Machine*. Journal of Physics: Conference Series, 971(1), 012056. <https://doi.org/10.1088/1742-6596/971/1/012056>
4. Aprilia, E. R., & Hardi, S. M. (2021). *Implementation of TF-IDF and Support Vector Machine for sentiment classification of e-commerce product reviews*. Jurnal RESTI, 5(2), 292–298. <https://doi.org/10.29207/resti.v5i2.2885>
5. Nugroho, A., Wibowo, F. W., & Putra, H. A. (2020). *Comparison of Sentiment Analysis using Naïve Bayes and Support Vector Machine Algorithms on Indonesian Online Shop Reviews*. Jurnal Ilmiah Teknik Industri, 19(1), 11–18. <https://doi.org/10.23917/jiti.v19i1.9587>
6. Purnamasari, L., Handayani, M. T., & Utomo, A. B. (2021). Analisis Sentimen Review Pengguna Google Play Store Menggunakan Text Mining dan Naive Bayes. Jurnal Informatika dan Rekayasa Perangkat Lunak, 9(1), 1–8.
7. K. S. Jones, “A statistical interpretation of term specificity and its application in retrieval,” *Journal of Documentation*, vol. 28, no. 1
8. Saurav Pradha, Malka N. Halgamuge, dan Nguyen T. Q. Vinh, “Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data,” in *Proc. of the 11th International Conference on Knowledge and Systems Engineering (KSE)*, Vietnam, 2019, pp. 103–108.
9. E. R. Aprilia dan S. M. Hardi, “Implementation of TF-IDF and Support Vector Machine for sentiment classification of e-commerce product reviews,” Jurnal RESTI, vol. 5, no. 2, pp. 292–298, 2021. <https://doi.org/10.29207/resti.v5i2.2885>
10. Nugroho, A., Wibowo, F. W., & Putra, H. A. (2020). *Comparison of Sentiment Analysis using Naïve Bayes and Support Vector Machine Algorithms on Indonesian Online Shop Reviews*. Jurnal Ilmiah Teknik Industri, 19(1), 11–18.